

Determination of Genomic Breed Composition using ADMIXTURE Software

Komal N. Patel^{1*}, Ashish C. Patel², Heena Shah³, Nilesh G. Nayee³, Dharamshi N. Rank²

ABSTRACT

Information from a reference population of animals with accurate breeding values is used to predict the genetic merit of a test population of animals for which such accurate phenotypic information is not available. For crosses, these procedures may be used to predict the proportion of genes of one breed in crossbred cattle when sets of genotypes from purebred animals are available as a reference. The genomic breed composition in 6326 HF crossbred animals (as a test population) performing in Gujarat was determined from SNPs (single nucleotide polymorphism) genotyped by INDUSCHIP array using ADMIXTURE software. The genotypes of 651 purebred indigenous and 120 exotic animals were used as reference genotype for determination of genomic breed composition. Unsupervised clustering analysis using ADMIXTURE software was carried out to infer ancestry ratios for 250 purebred animals using 14718 pruned INDUSCHIP SNPs. It showed that all the *B. indicus* animals were assigned to their respective breeds. A supervised clustering using 14718 INDUSCHIP variants was performed to infer breed proportions in 6326 crossbred animals with keeping *K* value as 9. As several indigenous breeds show evidence of admixture, the screening and selection for admixture in candidate bulls for breed purity can improve breeding programs. The average proportion of exotic inheritance was 55% in crossbred population. Hence, SNP genotyping gives opportunity to determine the exact breed proportions in crossbred animals and will be useful in regulation of exotic inheritance in the crossbred population for the optimum productivity of the animals reared by farmers.

Key words: ADMIXTURE analysis, Exotic inheritance, Genomic breed composition, HF crossbreds, SNP genotyping

Ind J Vet Sci and Biotech (2022): 10.48165/ijvsbt.18.5.20

INTRODUCTION

The indigenous cattle breeds are known for their remarkable power of heat tolerance and disease resistance in hot climate as compared to exotic breeds. Use of crossbreeding can be an effective tool for replacing non-descript animals with high yielding crossbred cattle. For this purpose, Holstein Friesian and Jersey inheritance with non-descript animals should be maintained around 50.0-62.5% level for better production performance (Wakchaure *et al.*, 2015). The breeding policy of Gujarat state recommends that the exotic inheritance in crossbreds shall be maintained around 50% and resource-rich farmers residing in favourable environmental areas shall be permitted to raise exotic inheritance up to 75% (Livestock Breeding Policy, Govt. of Gujarat, 2018). In the field, some indigenous breeds are often admixed due to breeding and rearing patterns of rural livestock owners (Gajjar *et al.*, 2018). Lack of pedigree information makes it difficult to determine the proportion of *B. taurus* ancestry when selecting bull mothers to produce crossbred bulls. Also, genotyping data can accurately estimate the proportional ancestry of *B. indicus* animals in selection candidates for future pure breeding programs, as well as the breed composition in crossbreds.

Using genomic data to determine breed composition is a novel idea and is superior to using pedigree records as it was shown to be more accurate whilst not being prone to missing, inaccurate or incomplete records (Gobena *et al.*, 2018). Using INDUSCHIP it is feasible to determine the

¹Polytechnic in Animal Husbandry, Kamdhenu University, Rajpur (Nava), Himmatnagar, Sabarkantha, Gujarat-383010, India

²Department of Animal Genetics and Breeding, College of Veterinary Science and Animal Husbandry, Kamdhenu University, Anand, Gujarat-388001, India

³Animal Breeding Group, National Dairy Development Board (NDDB), Anand, India-388001

Corresponding Author: Komal N. Patel, Polytechnic in Animal Husbandry, Kamdhenu University, Rajpur (Nava), Himmatnagar, Sabarkantha, Gujarat-383010, India, e-mail: komalpatel8276@gmail.com,

How to cite this article: Patel, K. N., Patel, A. C., Shah, H., Nayee, N. G., & Rank, D. N. (2022). Determination of Genomic Breed Composition using ADMIXTURE Software. *Ind J Vet Sci and Biotech.* 18(5), 100-103.

Source of support: None

Conflict of interest: None

Submitted: 08/09/2022 **Accepted:** 30/10/2022 **Published:** 10/11/2022

breed purity of major Indian cattle breeds and determine proportions of various breeds in crossbreds. Admixture is the presence of multiple genetically distinct subgroups within a population (Wang *et al.*, 2005). Information from a reference population of animals with accurate breeding values is used to predict the genetic merit of a test population of animals for which such accurate phenotypic information is not available. For crosses, these procedures may be used to predict the proportion of genes of one breed in crossbred cattle when sets of genotypes from purebred animals are available as a

reference. Therefore, the present study was undertaken to determine the proportions of exotic inheritance in farmer's crossbred animals using ADMIXTURE software.

MATERIALS AND METHODS

The 777K HD genotype data of 120 exotic animals of two *Bos taurus* breeds (Holstein and Jersey) were retrieved from WIDDE (<http://widde.toulouse.inra.fr/widde/>) open source. The 777K HD chip genotype data of 651 animals from 14 indigenous breeds generated earlier at NDDDB were included in the analysis. Further, the INDUSCHIP 49496 SNPs genotype data of 6326 crossbred animals as test animals were also used in the analysis.

Quality Control

The 777K HD genotype data of 120 exotic animals (Holstein and Jersey) and 651 animals from 14 indigenous breeds were merged (771 animals) and variants from INDUSCHIP (49496) were extracted in the PLINK v.1.9 software and R-program. Alleles with MAF (minor allele frequency) <5% and alleles not in Hardy-Weinberg equilibrium (HWE p value <0.00001) were filtered out for further analysis. Animals with more than 10% missing genotypes (mind <0.1) and variants with 10% missing genotyping rate were excluded thus, keeping only SNPs with a 90% genotyping rate (geno <0.1). This data were thinned using thin filter (thin 0.3) by retaining each variant with probability of 0.3 which finally kept only 14718 variants and 547 cattle for further analysis.

ADMIXTURE Analysis

Unsupervised clustering analysis using ADMIXTURE software (Alexander *et al.*, 2009) was carried out to infer ancestry ratios for 250 (out of 547) purebred animals using 14718 pruned INDUSCHIP SNPs. The obtained Q values (the ancestry coefficients) were plotted breed wise to show proportion of ancestry from different breeds in each individual animal. The pruned dataset was processed for ancestry and admixture analysis using the model based approach for K value = 9 (where, K value parameter describes the number of subpopulations that make up the total population) in ADMIXTURE program. Based on admixture analysis results,

breed composition of crossbred cattle was determined based on Q value (Alexander *et al.*, 2009)). The supervised ADMIXTURE analysis was carried out for 6326 genotyped crossbred animals.

RESULTS AND DISCUSSION

A total of 49496 INDUSCHIP SNPs were extracted from HD genotype data of 771 animals (120 animals from 2 exotic breeds, HF and Jersey plus 651 animals from 14 indigenous breeds) in the PLINK v.1.9 software and R-program. SNPs with minor allele frequency (MAF) <5% and alleles not in Hardy-Weinberg equilibrium (HWE p value <0.00001) were filtered out. Thus, 506 SNPs and 16 animals were excluded from the analysis. The 48990 INDUSCHIP variants and 755 animals that passed QC filter were available for further analysis. These data were thinned down using thin filter (thin 0.3) which removed 34272 variants at random which finally kept only 14718 variants for further analysis. Among 14 cattle breeds, 9 relevant breeds, *viz.*, HF, Jersey, Gir, Sahiwal, Red Sindhi, Kankrej, Tharparkar, Amritmahal and Kangayam were included in the study. By keeping 9 breeds, a total 547 (out of 755) animals were available for further admixture analysis.

ADMIXTURE Analysis

Figure 1 depicts results of ADMIXTURE analysis with $K=9$ for animals of seven *Bos indicus* and two *Bos taurus* breeds, which was found to be optimal to separate the breeds as distinct populations. Unsupervised clustering analysis showed that all the *B. indicus* animals were assigned to their respective breeds. In present study all breeds got separated well according to their purity. However, Kankrej showed admixture with Gir, Red Sindhi, Sahiwal and Tharparkar, but no any admixture from two draft breeds, *viz.*, Amritmahal and Kangayam. Red Sindhi and Sahiwal showed considerably less admixture from other breeds. Gir, Tharparkar, Amritmahal and Kangayam breeds showed highest purity among indigenous breeds. Both the exotic breeds also showed negligible admixture from either of the breeds. As several indigenous breeds show evidence of admixture, screening and selection for admixture in candidate bulls for breed purity can improve breeding programs. Gajjar *et al.* (2018) also performed Admixture

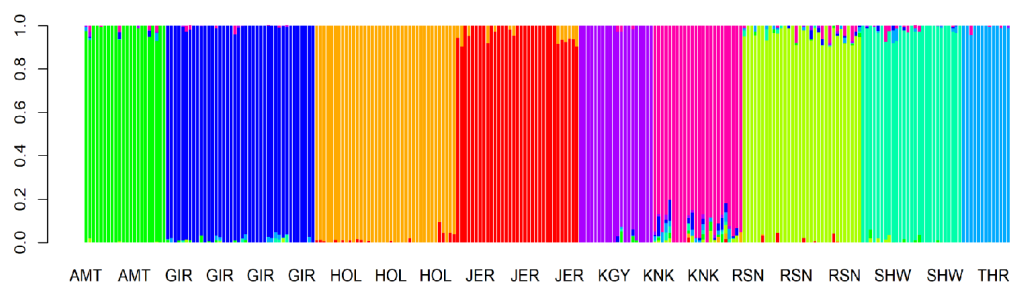


Fig. 1: Graph on ADMIXTURE analysis with $K=9$

(Vertical bars represent individuals. Breeds are indicated on the X-axis and ancestry ratios on Y-axis)

(AMT = Amritmahal, GIR = Gir, HOL = Holstein Friesian, JER = Jersey, KGY = Kangayam, KNK = Kankrej, RSN = Red Sindhi, SHW = Sahiwal, THR = Tharparkar)

analysis with $K = 12$ and suggested that indigenous breeds could be traced to distinct populations, except Hariana, Rath, Khillar and Hallikar. They found an evidence of migration from Gir to Deoni, however; no evidence of migration was noted to /or from *B. indicus* breeds of India to other global breeds.

Out of 547 cattle, only 250 purebred individuals were selected as reference population based on their purity with the Q value of >0.99 for all breeds, except >0.98 for Holstein, >0.95 for Gir and >0.80 for Kankrej. Unsupervised clustering analysis showed that all the *B. indicus* animals were assigned to their respective breeds.

A supervised clustering using 14718 INDUSCHIP variants was performed to infer breed proportions in 6326 crossbred animals keeping K value as 9 (K value: parameter describing the number of subpopulations that make up the total population). Based on admixture analysis results, breed composition of crossbred cattle was determined from Q value (the ancestry coefficients), (Alexander *et al.*, 2009). Crossbreds could well be identified, in accordance with known patterns of crossbreeding. In present study, an average breed proportions observed in HF crossbred cattle are shown in Table 1.

Table 1: Average genomic breed composition (GBC) in crossbred animals determined by SNP genotyping

Inheritance	Breed Name	GBC (%)		
		Minimum	Maximum	Average
Exotic	Holstein (HOL)	0	100	46
	Jersey (JER)	0	66	9
Indigenous	Gir (GIR)	0	100	7
	Sahiwal (SHW)	0	98	14
	Kankrej (KNK)	0	75	14
	Others (Red Sindhi + Amritmahal + Kangayam + Tharparkar)	0	37	10

The farmers in the target districts do not rear Gir, but rear Kankrej or non-descript animals, while the crossbred sires used for AI were mostly HFXSahiwal or HFXKankrej cross and rarely HFXGir cross. Hence, the average inheritance of Gir in the crossbred population was just 7% only. The average proportion of exotic inheritance from HF and Jersey was 55% in crossbred animals. However, average contribution from Jersey breed was much less, only 9% as compared to HF in the crossbred population.

Based on genomic breed proportions, 4927 (out of total 6326) genotyped crossbred animals were categorized into six genetic grades $<40\%$ ($n=156$), $40-50\%$ ($n=1202$), $50-60\%$ ($n=2202$), $60-70\%$ ($n=1031$), $70-80\%$ ($n=246$) and $>80\%$ ($n=90$). Among all the crossbreds, the crossbreds with 50 to 60 % exotic inheritance were found to be significantly ($p<0.01$) higher than other genetic grades. The present

finding indicated that the majority of crossbred population found in India may be with 50-60% exotic inheritance only, followed by 40-50%, 60-70%, 70-80% and less than 40%. Only 2 % of tester population was found with more than 80 % of exotic inheritance. In accordance with present study Gebrehiwot *et al.* (2021) estimated breed proportions of crossbred animals from Kenya, Ugandan, Ethiopian, Tanzanian and Senegal using the 38 k and 713 k SNP datasets which shown an average dairy proportion of 66% (± 0.199), 58% (± 0.190), 68% (± 0.168), 69% (± 0.169), and 51% (± 0.179), respectively.

Admixture of exotic breeds into native cattle breeds in 829 animals from 14 indigenous breeds (*Bos indicus*) was investigated by Gajjar *et al.* (2018) using 777K HD chip (Illumina Inc., San Diego, CA) genotype data on 20 each from Danish Jersey and Nordic Holstein breeds and 777K HD chip genotype data on 2 *Bos taurus* samples and 50K chip (Illumin Inc., San Diego, CA) genotype data on 29 cattle breeds. All indigenous breeds, except Hariana, Rath, Khillar and Hallikar, suggested a distant origins in ADMIXTURE analysis. They found stable proportions of four populations each in Khillar and Hallikar, which suggested a distant origin by admixture for these breeds, while Hariana and Rath showed signs of recent admixture. No evidence for migration to/or from indigenous breeds of India to other cattle breeds was found. A set of 500 ancestry informative markers for Gir, Sahiwal, Kankrej, Red Sindhi, Holstein Friesian and Jersey breeds was identified and found suitable for tracing ancestries of these purebreds as well as their crossbreds.

In another attempt, supervised clustering was carried out by Nayee *et al.* (2018) to infer ancestry ratio in crossbred animals keeping $K = 6$ (Sahiwal, Gir, Kankrej and Red Sindhi and their crosses with two *Bos taurus* breeds, Holstein and Jersey) using Gene-Seek 70K INDICUS chip to know the best breed composition. Unsupervised clustering analysis using ADMIXTURE software (Alexander *et al.*, 2009) showed that the majority of the *B. indicus* animals were assigned to their respective breeds. A supervised model-based ancestry estimation using 777K chip and $K = 6$ for crossbred animals of four *B. indicus* and two *B. taurus* breeds evidenced that some individuals from *B. indicus* breeds had mixed ancestry. They also observed that degree of admixture varied from individual to individual.

CONCLUSIONS

Most of the time farmers are unaware of the deep pedigree of their crossbred animals. Even if the pedigree is available the pedigree based breed proportions may not reflect exact proportions of exotic inheritance due to recombination of genes. SNP genotyping gives opportunity to determine the exact breed proportions in crossbred animals and will be useful in regulation of exotic inheritance in the crossbred population for the optimum productivity of the animals reared by farmers. The present study is the first of its kind in India on estimation of genomic breed composition without



using pedigree information through SNP genotyping in crossbred animals reared by farmers. SNP genotyping will be useful in the regulation of exotic inheritance in the crossbred animals for the optimum productivity in the field.

ACKNOWLEDGEMENTS

We are thankful to National Dairy Development Board for providing us the genotypic and phenotypic data on crossbred cattle for analysis.

REFERENCES

- Alexander, D.H., Novembre, J. & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19 (9), 1655-1664.
- Gajjar, S.G., Guldbbrandtsen, B., Nayee, N., Sudhakar, A., Trivedi, K., Lund, M.S., & Sahana, G. (2018). Genetic composition, divergence, admixture and use of ancestry informative markers in indigenous breeds of cattle in India. In: *Proceedings of the 11th World Congress on Genetics Applied to Livestock Production*, February 2018, pp. 11-16.
- Gebrehiwot, N. Z., Strucken, E. M., Marshall, K., Aliloo, H., & Gibson, J. P. (2021). SNP panels for the estimation of dairy breed proportion and parentage assignment in African crossbred dairy cattle. *Genetics Selection Evolution*, 53 (1), 1-18.
- Gobena, M., Elzo, M.A., & Mateescu, R.G. (2018). Population structure and genomic breed composition in an Angus-Brahman crossbred cattle population. *Frontiers in Genetics*, 9, 1-10.
- Livestock Breeding Policy, Government of Gujarat. (2018). Livestock (Bovine) Breeding Policy Notification, Retrieved from https://www.dairyknowledge.in/sites/default/files/pdfs/breeding_policies/Gujarat-State-Breeding-Policy.pdf
- Nayee, N., Sahana, G., Gajjar, S., Sudhakar, A., Trivedi, K., Lund, M.S., & Guldbbrandtsen, B. (2018). Suitability of existing commercial single nucleotide polymorphism chips for genomic studies in *Bos indicus* cattle breeds and their *Bos taurus* crosses. *Journal of Animal Breeding and Genetics*, 135(6), 432-441.
- Wakchaure, R., Ganguly, S., Para, P.A., Praveen, P.K., Kumar, A. & Sharma, S. (2015). Development of crossbred cattle in India: A review. *International Journal of Emerging Technology and Advanced Engineering*, 5, 75-77.
- Wang, W.Y., Barratt, B.J., Clayton, D. G. & Todd, J.A. (2005). Genome-wide association studies: Theoretical and practical concerns. *Nature Reviews Genetics*, 6 (2), 109-118.
- WIDDE (Web-Interfaced next generation Database dedicated to genetic Diversity Exploration). Retrieved from <http://widde.toulouse.inra.fr/widde>